

TITLE OF THE INVENTION  
CONTROL METHOD AND DEVICE OF JITTER BUFFER

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a control method of a jitter buffer that temporarily stores communication packets in the VOIP (Voice Over Internet Protocol).

2. Description of the Related Art

In recent years, the internet telephone has remarkably been spreading, and the sound quality thereof has been improved to such a level that practical inconveniences can hardly be found. And, the means to attain this level has adopted a method, which temporarily stores received packets in a jitter buffer, and thereby absorbs jitters accompanied with the packets. The extent of jitters in the packet transmission through the internet is about 100 ms, however the frame length in the packet transmission conforming to the International Standard G.729 is 10 ms. Accordingly, the jitter buffer needs a capacity to contain the packets of 10 frames at the minimum. However in the VOIP, if the capacity of the jitter buffer is too large, it will produce a delay time proportional to the capacity of the jitter buffer, after the counterpart finishes talking. This delay will interrupt the conversation, and further make produced echoes stand out, which deteriorates the sound quality of talks.

In order to solve this problem, a conventional technique, for example, sets a "disposal starting threshold" and a "disposal ending threshold" in the jitter buffer, executes the

disposal processing of the packets according to the contents of voice data to thereby restrict the capacity of the jitter buffer, and inserts the "voice data of minute noises" to prevent interrupting voices and sounds (patent document 1: the Japanese Published Unexamined Patent Application No.274829/2001).

Or, another conventional technique measures an available residual quantity of a reception buffer (jitter buffer), determines a clock frequency for setting the read timing in order that the available residual quantity always be within a predetermined range, and thereby restricts the capacity of the reception buffer (patent document 2: the Japanese Published Unexamined Patent Application No.261613/1997).

However, the technique disclosed in the patent document 1 determines to or not to apply the disposal processing according to the contents of the voice data; accordingly, if there occurs a burst of jitters, it will lead to extinguishing a large quantity of packets, which produces distortions on reproduced voices. And, if such a processing is executed to a silent interval, such distortions will not be produced; however a calculation processing of the sound pressure becomes necessary, and the load to the software and hardware becomes increased accordingly.

Or, the technique disclosed in the patent document 2 determines the clock frequency for setting the reading speed by measuring the available residual quantity of the reception buffer, and controls the capacity of the reception buffer. However, the document does not disclose as to what kind of countermeasures can be taken, when there occurs a burst of jitters and the available residual quantity is nonexistent.

## SUMMARY OF THE INVENTION

This invention has been made in view of the above problems, and it provides a method and device for controlling a jitter buffer, which avoids the delay of packets by adding and deleting the packets to enhance the sound quality of talks. The method further lessens the loss of communication packets by fluctuating the reproduced clock frequency to thereby reduce the distortions of voices.

In order to accomplish the above problems, according to one aspect of the invention, the method of controlling a jitter buffer sets a packet delete area, a packet add area, and a clock control area inside a FIFO that forms the jitter buffer. The method controls to delete the packets when a stored packet quantity  $T_j$  of the FIFO is within the packet delete area, to add the packets when the stored packet quantity  $T_j$  is within the packet add area, and to raise or lower the clock frequency for reading the packets when the stored packet quantity  $T_j$  is within the clock control area, in which the clock control area is set between the packet add area and the packet delete area.

## BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a conceptual chart explaining the method of controlling a jitter buffer in the first embodiment of the invention;

Fig. 2 is an explanatory chart for the operation of the jitter buffer in the first embodiment of the invention;

Fig. 3 is a block diagram of a control device for the jitter buffer in the first embodiment of the invention;

Fig. 4 illustrates state transition diagrams in the first

embodiment of the invention; and

Fig. 5 illustrates an example of controlling the clock frequency in the second embodiment of the invention.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The preferred embodiments of the invention will be described in detail with reference to the accompanying drawings. Fig. 1 is a conceptual chart explaining the method of controlling the jitter buffer in the first embodiment of the invention. Generally, a jitter buffer 101 is configured with a FIFO (First In First Out), and the voice packets written in from the input terminal are located to the trailing one of the voice packets having been arranged from the output terminal, which are sequentially read out from the output terminal. A packet stored quantity surveillance described later detects the trailing packet of a voice packet string, and detects the stored packet quantity  $T_j$  at that moment. In this embodiment, the method uses this stored packet quantity  $T_j$  as a guidepost in controlling the jitter buffer 101, and sets two areas inside the jitter buffer.

One of the two areas is a buffer control area 102, which includes a packet delete area 104 and a packet add area 105. This area 102 sets the upper limit  $T_1$  of the packet add area 105 being the practical minimum size of the jitter buffer 101, and the lower limit  $T_4$  of the packet delete area 104 being the maximum size thereof.

The upper limit  $T_1$  of the packet add area is the minimum buffer size for absorbing jitters. If the quantity of packets becomes lower than this value, it will increase the possibility

of producing missing voice packets, while the jitters cannot be absorbed. Therefore, it is preferable that this value T1 is preset with reference to actual measurements and the like.

On the other hand, the lower limit T4 of the packet delete area is related to the delay of the voice signal in the VOIP. To set this value larger will produce harmful effects such as echoes created by the delay of packets, which deteriorates the sound quality of talks. Therefore, it is also preferable in the same manner as above that this value T4 is preset with reference to actual measurements and the like. The upper limit T5 of the packet delete area is a means not to increase any more the voice packets stored in the jitter buffer 101. Since the packets inputted to exceed this value are to be deleted, when the jitter buffer 101 is built up inside a RAM, this value T5 is equivalent to the physical maximum size thereof.

Another one area in the jitter buffer is a clock control area 103, which is allocated between the levels T2 and T3. In the same manner as above, the levels T2 and T3 are preset on the basis of actual measurements and the like. When the stored packet quantity Tj becomes lower than the level T2, the reading clock frequency is controlled into a lower frequency. When the stored packet quantity Tj exceeds the level T3, the reading clock frequency is controlled into a higher frequency.

The control of the reading clock frequency is executed to have a hysteresis characteristic. That is, the moment when the currently stored packet quantity Tj becomes lower than the level T2, even if it restores a level exceeding T2 immediately after that moment, the clock frequency will not be changed. In the same manner, the moment when the currently stored packet

quantity  $T_j$  becomes higher than the level  $T_3$ , even if it returns to a level lower than  $T_3$  immediately after that moment, the clock frequency will not be changed. The reason for this control lies in that, since the stored packet quantity  $T_j$  fluctuates from moment to moment, to change the clock frequency each time leads to producing distortions on reproduced sounds in some cases.

According to this embodiment, the clock control area is set between the packet add area 105 and the packet delete area 104 of the buffer control area. The reason for this setting is based on a consideration that, since the change of the clock frequency is gradual and the fluctuation thereof cannot be set to a large level, it is very difficult to adapt a situation that needs a quick decrease of the buffer size, immediately after when there occurs a sporadic jitter burst.

As the result, when the stored packet quantity  $T_j$  at the present moment is about to exceed the lower limit  $T_4$  of the packet delete area 104, the voice packets are deleted, which maintains the maximum size of the buffer control area. Also, when the stored packet quantity  $T_j$  becomes lower than the upper limit  $T_1$  of the packet add area, the voice packets are refilled, which maintains the minimum size of the buffer control area.

When the stored packet quantity  $T_j$  exceeds the lower limit  $T_4$  of the packet delete area 104, and the packets are deleted, if the packets are located inside a silent interval, it is conceivable that there is not any influence on the reproduced sound quality. But, if the voice packets being located inside an interval with sounds are deleted at a burst, conceivably it will give a significant influence on the reproduced sound quality. In order to minimize this influence, it is necessary

to expand the difference between the levels T3 and T4, or to increase the extent of raising the clock frequency. That is, the method raises the clock frequency a little early to absorb the jitters, and thereby reduces the number of the voice packets to be deleted.

And, there are several methods for refilling the voice packets, when the stored packet quantity  $T_j$  becomes lower than the upper limit T1 of the packet add area. One of the most general methods is to output the same packet as the previous one. Also in this case, in the same manner as above, if the packets inside an interval with sounds are added at a burst, conceivably it will give an influence on the reproduced sound quality to no small extent. In order to minimize this influence, in the same manner as above, it is necessary to expand the difference between the levels T2 and T1, or to increase the extent of lowering the clock frequency. That is, the method lowers the clock frequency a little early to reduce the number of the voice packets to be added.

However, even when the stored packet quantity  $T_j$  is reduced lower than the upper limit T1 of the packet add area, the voice packets are present in the buffer. Therefore, it is not necessarily needed to add the packets, or even if it is needed, it is possible to add the packets scatteredly with received voice packets mixed.

Here, it is possible to set the level of the packet add area to  $T_1 = T_0$ . However, this case will have to continue adding the packets at the moment when the stored packet quantity  $T_j$  reached this level.

Fig. 4 illustrates state transition diagrams in the first

embodiment, and Fig. 4(A) illustrates a transition state of the clock frequency. The transition state of the clock frequency possesses an ascending state 402 that changes from CLK0 into CLK1, and a descending state 401 that changes from CLK1 into CLK0. When the clock frequency is in the descending state at the beginning, and the stored packet quantity  $T_j$  at that moment exceeds the level T3, the state changes into the state to ascend the clock frequency, which will accelerate the processing of the packets stored in the buffer. In reverse, when the clock frequency is in the ascending state at the beginning, and the stored packet quantity  $T_j$  at that moment falls below the level T2, the state changes into the state to descend the clock frequency, which will prevent the situation that the voice packets in the jitter buffer 101 are insufficient.

The transition diagram in Fig. 4(B) illustrates the transition state relating to the addition/deletion of the voice packets. As shown in the drawing, the state transition diagram possesses a delete state 403 of the voice packets, an add state 405 of the voice packets, and a state 404 that neither adds nor deletes the voice packets. The initial state is assumed to be the state 404 that neither adds nor deletes, and as the stored packet quantity  $T_j$  exceeds the lower limit T4 of the packet delete area 104, the state shifts into the delete state 403 to delete the voice packets, which will maintain the maximum size of the buffer control area. As the stored packet quantity  $T_j$  falls below the T4, the state returns to the initial state 404.

And, as the stored packet quantity  $T_j$  exceeds the upper limit T5 of the packet delete area 104, the state remains at the delete state 403. On the other hand, when the state is the



initial, as the stored packet quantity  $T_j$  falls below the upper limit  $T_1$  of the packet add area 105, the state shifts into the add state 405 of the voice packets, which will maintain the minimum size of the buffer control area. As the  $T_j$  exceeds the  $T_1$  at this moment, the state returns to the initial state 404. And, as the stored packet quantity  $T_j$  is about to fall below the lower limit  $T_0$  of the packet add area 105, the state will remain at the add state 405 of the voice packets.

Even when the stored packet quantity  $T_j$  is about to fall below the lower limit  $T_0$  of the packet add area, there remain the packets in the buffer, and it is not necessarily needed to add the packets. However, to add the packets sparsely with the received packets mixed will prevent intermittence of voices without increasing distortions of reproduced voices.

Now, the operation of the jitter buffer 101 in the first embodiment will be described concretely. Fig. 2(A) illustrates a concrete example of the quantity of jitter, which shows the quantity of jitter against the time base in unit of millisecond. The jitter is defined as an arrival time interval with the previous packet, and is given by a relative value. As shown in the drawing, the jitters of voice packets frequently occur at a burst through the network.

Fig. 2B) illustrates the stored quantity of voice packets into the buffer, which shows the packet stored quantity in unit of packet against the time base. In this illustration, the stored packet quantity  $T_j$  increases accompanying the jitter burst; it exceeds the level  $T_3$  at time  $t_1$ , and it exceeds the lower limit  $T_4$  of the packet delete area 104 at time  $t_2$ . When the jitter burst disappears to decrease the stored packet

quantity  $T_j$ , it falls below the level  $T_2$  at time  $t_4$ , and it falls below the upper limit  $T_1$  of the packet add area 105 at time  $t_5$ .

Fig. 2(C) shows a reproduced clock frequency against the time base. As shown in the drawing, the clock frequency is not changed discretely, but it is controlled continuously to vary smoothly. To control in this manner is to avoid that to change the clock frequency sharply creates distortions on reproduced sounds, which deteriorates the sound quality. In the example of the drawing, the change of the clock frequency is linear. The clock frequency falls up to time  $t_1$ , and after the stored packet quantity  $T_j$  exceeds the level  $T_3$  at time  $t_1$  (see Fig. 2(A)), the clock frequency rises from  $CLK_0$  to  $CLK_1$ , which increases the packet quantity read from the jitter buffer 101. This suppresses a sharp increase of the stored packet quantity, which reduces the number of voice packets to be deleted. Since the stored packet quantity  $T_j$  falls below the level  $T_2$  at time  $t_4$ , the clock frequency starts falling toward  $CLK_0$ . This delays the speed of reading the packets from the jitter buffer 101, and lowers the provability by which the voice packets enter the packet add area, which makes it possible to reduce the number of voice packets to be added, thus contributing to enhancing the sound quality.

Fig. 2(D) shows the number of the packets to be added/deleted against the time base, when there occur the jitter fluctuations as shown in Fig. 2(A), and the stored packet quantity  $T_j$  of the jitter buffer 101 changes as shown in Fig. 2(B). Since the stored packet quantity  $T_j$  exceeds the lower limit  $T_4$  of the packet delete area 104 at time  $t_2$  through time  $t_3$ , the packets are deleted. Since the stored packet quantity

Tj falls below the upper limit T1 of the packet add area 105 at time t5 through time t6, the packets are added.

Next, the control device for controlling the jitter buffer 101 of this embodiment will be described. Fig. 3 illustrates a circuit configuration of the control device for the jitter buffer 101 in this embodiment.

This control device is made up mainly with a jitter buffer 301 and a jitter buffer control circuit 302. The jitter buffer 301 is generally configured with a FIFO built up inside a memory (RAM), however it can employ a FIFO as hardware.

The jitter buffer control circuit 302 is configured with a buffer accumulation level surveillance 303 that monitors the stored packet quantity Tj of the voice packets accumulated in the jitter buffer 301, a VCO (voltage controlled oscillator) 304 that supplies to vary the reproduced clock frequency as required, a buffer control circuit 305 that controls the operations of the jitter buffer 301 and the peripheral circuits. In replacement for the VCO, a PWM (Pulse Width Modulator) can be used as well.

The control device provides a packet deletion circuit 306 and a packet addition circuit 307 in the pre-stage and post-stage of the jitter buffer 301. The packet addition circuit 307 adds a specified packet under the control of the buffer control circuit 305, when the stored packet quantity Tj falls below the upper limit T1 of the packet add area inside the buffer control area. The packet deletion circuit 306 deletes the voice packets under the control of the buffer control circuit 305, when the stored packet quantity Tj exceeds the lower limit T4 of the packet delete area 104.

The voice packets entering the jitter buffer 301 pass through the packet deletion circuit 306 to be accumulated in the jitter buffer 301, and then pass through the packet addition circuit 307 to be delivered thereafter to a decoder 308.

The decoder 308 accepts the packets outputted from the packet addition circuit 307, and sends the frames of the voice packets being the contents of the packets outputted to a D/A converter 309 based on the clock signal supplied from the VCO 304. The D/A converter 309 converts the voice digital data of the frames into the analog signals to be outputted to a speaker 310, and the speaker 310 emits the voices reproduced.

Second Embodiment:

Fig. 5 illustrates a method of controlling the clock frequency in the second embodiment of the invention. The other components of the jitter control method are the same as those in the first embodiment. The clock frequency is controlled linearly in the first embodiment, however in the second embodiment, the clock frequency is controlled exponentially, which is the difference of both. That is, the control is executed based on the following expression.

$$\text{CLK} = \text{CLK0} + (\text{CLK1} - \text{CLK0}) * (1 - \text{EXP}(-T/\text{Td}))$$

By setting the time constant  $\text{Td}$  in this expression smaller, instead of varying the clock frequency linearly at time  $t1$  through time  $t4$ , the method varies at time  $t1$  the clock frequency from the frequency of  $\text{CLK0}$  to that of  $\text{CLK1}$  in a very short time, and varies at time  $t4$  the clock frequency from the frequency of  $\text{CLK1}$  to that of  $\text{CLK0}$  very quickly. The method controls the clock frequency in this manner to follow a sharp change in the accumulated quantity of the voice packets, lowers the

provability by which the stored packet quantity  $T_j$  enters the packet delete area or the packet add area, and reduces the number of the packets to be deleted/added, thereby enhancing the sound quality of the reproduced voices.

The embodiments being thus described, the method of controlling the jitter buffer according to the invention provides the packet delete area and packet add area, the clock control area inside the FIFO that forms the jitter buffer. And, the method controls to delete the packets when the stored packet quantity  $T_j$  of the FIFO is within the packet delete area, to add the packets when the stored packet quantity  $T_j$  is within the packet add area, and to raise or lower the clock frequency for reading the packets when the stored packet quantity  $T_j$  is within the clock control area. Here, the clock control area is located between the packet add area and the packet delete area. Therefore, if there occurs a jitter burst, the method lowers the provability by which the packets accumulated in the FIFO enter the packet delete area or the packet add area, and lowers the provability to add/delete the packets, thus reducing distortions on the reproduced voices.